3D Videoconferencing System Using Spatial Scalability

M. Chavarrias, F. Pescador, F. Jaureguizar, E. Juarez, M.J. Garrido.

***Abstract*—An implementation of a real-time 3D videoconferencing system using the currently available technology is presented. This approach is based on the side by side spatial compression of the stereoscopic images. The encoder and the decoder have been implemented in a standard personal computer and a conventional 3D compatible TV has been used to present the frames. Moreover, the users without 3D technology can use the system because 2D compatibility mode has been implemented in the decoder. The performance results show that a conventional computer can be used for encoding/decoding audio and video streams and the delay in the transmission is lower than 200 ms.**

I. INTRODUCTION

Research and development in the area of videoconferencing from a local to a remote side has a long tradition. One of the most important objectives in these researches is the improvement of the realism and the filling of closeness.

Currently, 3D video is entering broad in the mass markets. Cinemas are being upgraded to 3D and available material is being converted from 2D to 3D. The technology is now maturated, providing excellent quality [1]. With more and more 3D cinemas and 3D movies being available, 3D also becomes increasingly interesting for other applications like home enter-tainment, mobile devices or videoconferencing systems [2].

Several researches in the implementation of videoconferencing and telepresence systems has been done during the last years [1][3][4]. All of them use the depth or disparity for adaptation to different displays by rendering or synthesizing virtual views. All these implementations require high computational systems for video encoding and special monitors for the presentation of the images, so nowadays they are not available for the consumer electronic market.

This paper presents a low cost 3D video conferencing system using devices that can be found currently in the market and computers without special features. This system use the technique based in the spatial compression of the right and left images [1] and the H.264 video compression standard for encoding the video sequence.

The rest of this paper is organized as follows: in section II, the system architecture is outlined; in section III, the performance results are explained; finally, section IV is devoted to the conclusion.

This work was supported by the Spanish Ministry of Science and Technology under grant TEC2009-14672-C02-01.

M. Chavarrias, F. Pescador, E. Juarez and M.J. Garrido are with the Electronic and Microelectronic Design Group (GDEM). ({m.chavarrias, pescador, ejuarez, matias}@sec.upm.es). F. Jaureguizar is with Image Processing Group (GTI) (fjn@gti.ssr.upm.es). Both groups at the Universidad Politecnica de Madrid, Spain.

II. PROPOSED ARCHITECTURE

This section describes the system specifications and implemented architecture using an example of the system configuration (capture, encoding, transmission and visualization).

A. System specifications

Unlike existing systems for 3D video conferencing a solution is proposed based on currently available technologies that allow implementing a video conferencing system using available low cost devices. This intends to employ a 3D compatible TV, a personal computer using a conventional H.264 encoder, an IP-STB (or the same personal computer used for encoding) and different spatial compression formats like side by side, top and bottom or checkerboard [4]. To meet the delay requirements needed in video conferencing applications (less than 200 ms), the encoding and decoding processes has been optimized eliminating the intermediate buffers that generate major delays in the transmission chain.

Additionally, for users without capacity to capture and display 3D images, both, the encoder and decoder can operate in 2D.

B. System architecture

Communications in a videoconferencing system must be bidirectional, thus the equipment available for transmission and reception should be similar. To simplify the explanation, only the architecture that allows one-way communications is explained. There are three different functional blocks in this architecture: the transmitter that includes the capturing and encoding devices, the receptor composed of decoding and display elements and the transmission medium. Fig 1 shows the block diagram of the system.

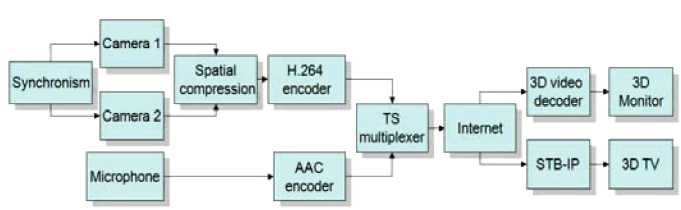

Fig. 1. System block diagram.

The transmitter system consists of two low-cost PAL resolution analog cameras with a synchronization input, this signal is provided by a synchronism source which indicates the frame rate and the same exactly instants of frame capture to the cameras. The output from both cameras is connected to two video capture cards [5] housed in the personal computer to be used as video encoder. A specific application has been developed to downscale both images and generates a frame in one of the three available formats [6] (top and bottom, side by side or checkerboard), rescales the resulting composed image to HD format and encoding the stream using H.264 video coding standard. The microphone information is compressed using AAC audio coding format in parallel with video coding. Both elementary streams can be sent directly through the local network using RTP/UDP protocol, or encapsulated in an MPEG-2 transport stream which is then sent using IP packets. The application has been implemented on an eight-core 3 GHz personal computer with 6 GB of RAM memory.

The transmission system is an IP network. A VPN between the two sites has been used to guarantee Quality of Service (QoS). This VPN allows bandwidths above 30 Mbps.

The receiver can be implemented using the encoder computer or a conventional IP-STB [7]. In both cases, the coded frames are received through the local network, and displayed on a TV with 3D presentation capabilities [8].

If the user only has one camera, the encoder can be configured to encode the information directly, without the need for spatial scaling. Moreover, the encoder has been designed to support cameras with different spatial resolutions because it is able to rescale the input frames from any size to the desire output resolution (normally 1080p). On the other hand, if the receiver is not capable of viewing 3D images and the personal computer is used as decoder, it can be configured to display only a one-eye image on screen.

III. RESULTS

Several tests have been done with the system described in section II. The most important results are related with the bitrate needed for a good Quality of Experience (QoE), the encoder computational load and the latency in the transmission.

The bitrate of the encoded sequences is one of the most important parameters in the subjective quality perceived by the users. Some subjective tests have been done with the implemented systems using different bitrates. The users remarked that 10 Mbps is the minimum bitrate needed to see the 3D effect using high definition frames (1080p) but 12 Mbps is more suitable for this application.

The QoE could be also affected by the spatial compression method employed. Subjective tests showed that exist a drop in QoE using checkerboard instead of Side by Side or Top & Bottom formats.

Two measures related with the computational load of the encoder have been done. An eight-core computer working at 3 GHz with 6 GB of RAM memory and a bitrate of 12 Mbps have been used. Firstly, the computational load of the encoder has been measured with different spatial compression formats and with two types of sequences: "low" movement sequences and continuous movement sequences. Table I shows the computational load in all the situations using 1080p resolution frames. This data demonstrate that a conventional personal computer can be used for this application.

TABLE I
COMPUTATIONAL LOAD NEEDED WITH DIFFERENT SPATIAL COMPRESSIONS.

Spatial Compression	Low Movement	Continuous Movement
Top & Bottom	43%	59%
Side by Side	47%	65%
Checkerboard	40%	55%

Secondly, the computational load needed to encode 2D and 3D sequences has been measured using the same sequences than in the previous experiments. Table II shows the percentage of CPU used in each case. The increase of the computational load from encoding 2D sequences to encoding 3D sequences in around 10%.

TABLE II
COMPUTATIONAL LOAD USED FOR 2D AND 3D ENCODING.

Configuration	Low Movement	Continuous Movement
2D	55%	67%
3D	65%	74%

Finally, the latency has been measured. A 1 Gbps network has been used to connect both sites. The latency is less than 200 ms that is enough for a videoconference service.

IV. CONCLUSIONS

A videoconferencing service has been presented using some low cost devices. This service uses different formats for spatial compression of stereoscopic video signal. These techniques allow using standard 3D monitors and 3D compatible TV sets. The proposed system is compatible with 2D contents. Up to the best of our knowledge, no other video conferencing systems based in this technology has been reported.

REFERENCES

[1] I. Feldman, W. Waizenegger, N. Atzpadin, O. Shreerm. "Real-Time Depth estimation for immersive 3D videoconferencing", Proc. 3DTV-Conference, pp. 1-4, Tampere, Finland, Jun. 2010.

[2] I. Andorko, P. Corcoran, P. Bigioi, "Hardware implementation of a real-time 3D video acquisition system". Proc. Int. Conf. on Optimization of Electrical and Electronic Equipment, , pp. 920-925, Brasov, Romania, May 2010.

[3] P. Merkle, K. Müller and T. Wiegand, "3D Video: Acquisition, Coding, and Display". IEEE Trans. on Consumer Electronics, pp. 946 – 950, Vol. 56, No.2, May 2010.

[4] Atzpadin, P. Kauff, O. Schreer, "Stereo Analysis by Hybrid Recursive Matching for Real-Time Immersive Video Conferencing", IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Immersive Telecommunications, pp. 321-334, Vol. 14, No. 3, Jan 2004.

[5] DeckLink HD Extreme 3D capture card. Blackmagic design. http://www.blackmagic-design.com/products/decklinkhdextreme/.

[6] Main Concept home page. 3D Solutions. http://www.mainconcept.com/products/sdks/3d.html.

[7] Amino HD STB. AmiNET130 User guide. http://www.teracue.com/dirlist/download/03.User_Guide/02.Decoding-STB-Player/AmiNet/AmiNET130%20User%20Guide.pdf.

[8] Samsung 3D LED TV. http://www.samsung.com/uk/consumer/tv-audio-video/television/led-tv/UE40C7000WKXXU/index.idx.